International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 5 Issue 4, May-June 2021 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Fashion AI Literature

Ashish Jobson[1], Dr. Kamalraj R[2]

[1]MCA, [2]Associate Professor,
[1,2]Department School of CS and IT, Jain University, Bangalore, Karnataka, India

ABSTRACT

We concentrate on the task of Fashion AI, which entails creating images that are multimodal in terms of semantics. Previous research has attempted to use several class-specific generators, which limits its application to datasets with a limited number of classes. Instead, we suggest a new Group Decreasing Network (GroupDNet), which takes advantage in the generator of group convolutions & gradually reduces the percentages of the groups decoder's convolutions. As a result, GroupDNet has a lot of influence over converting semantic labels to natural images and can produce plausible high-quality results for datasets with a lot of groups. Experiments on a variety of difficult datasets show that GroupDNet outperforms other algorithms in the SMIS mission. We also demonstrate that GroupDNet can perform a variety of interesting synthesis tasks.

***KEYWORDS:** Multi-Layer Neural Network, Machine Learning*

***How to cite this paper**:* Ashish Jobson | Dr. Kamalraj R "Fashion AI Literature" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-5 | Issue-4, June 2021, pp.755-757, URL: www.ijtsrd.com/papers/ijtsrd42378.pdf

IJTSRD42378

I. INTRODUCTION

Fashion AI, which has a wide variety of real-world uses and draws a lot of interest because it converts semantic marks to natural images. In recent years, convolutional neural networks have been used to effectively complete object detection, object recognition, image segmentation, and texture synthesis. By itself, it's a one-to-many mapping challenge. A single semantic symbol may be associated with a wide number of different natural images. Using a variational auto-encoder, inserting noise during preparation, creating several sub networks, and using instance-level feature embeddings, among other methods, have been used in previous studies. While these approaches have made considerable strides in terms of image quality and execution, we take it a step further by working on a complex multi-modal image synthesis task that allows us to have greater control over the performance. While this has resulted in a significant improvement over the previous generation of features, these networks are learned in a fully supervised manner on vast volumes of data, requiring expensive and time-consuming annotation. Features learned on one dataset can be applied to another, but not all datasets are created equal, so features learned on Image Net will not perform as well on data from other datasets. Under an increasing number of classes, however, this type of approach quickly degrades in efficiency, increases training time linearly, and consumes computational resources.

II. METHODOLOGY

We provide comprehensive network designs for each dataset in this section. The discriminator's architecture is shown.

1. The discriminator's design remains consistent across datasets.
2. For various datasets, the encoder architecture is shown.
3. Depicts the Deep Fashion decoder architectures.
4. The design of the decoder for the ADE20K is seen. Since ADE20K has so many groups, we reduce the number of channels in each category to avoid excessive GPU consumption.

The total network capacity decreases in this situation, which we believe is not beneficial to the performance. Therefore, we add a few more convolutional layers to the mix, increasing network capacity; as a result, the ADE20K decoder varies from the other two datasets in terms of construction.

Training details

All experiments are educated for 100 epochs on Deep Fashion, with learning speeds for both the generator and discriminator remaining constant for the first 60 epochs before linearly decaying to zero in the last 40.

- The batch size for Deep Fashion is 32, while the batch size for ADE20K is 16. This is due to the vast number of channels needed to meet the specifications of sufficient capacity for the 150 classes.
- Glorot initialization is used to set the network weights after SPADE.

Strategy for choosing a group number

It's difficult to come up with an algorithmic approach to calculate the diminishing numbers when GPU memory capacity, batch size, and the number of variables is all restricted. To create the group numbers, however, we followed two rules:

1. In the first few layers of the decoder, the numbers decrease dramatically, lowering the computational cost significantly;
2. The earlier layer's group number is either the same as or twice as the next layers.

III. Working Flow

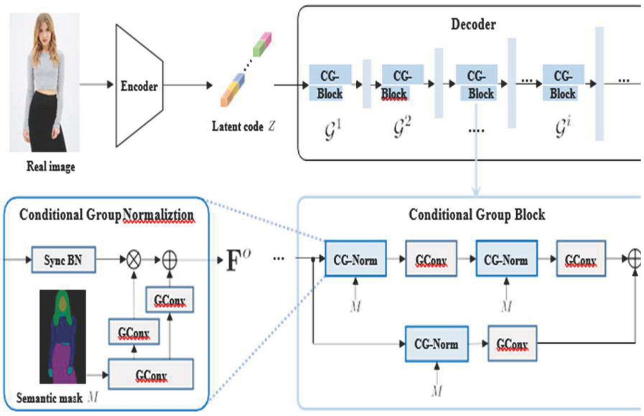

Figure 1: Architecture of our generator method.

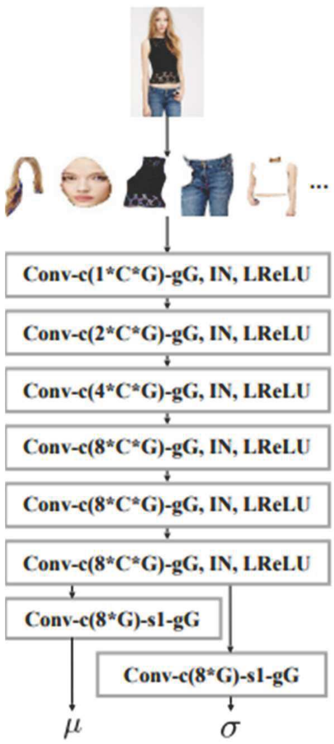

Figure 2: Functional Architecture

IV. RESULTS AND DISCUSSION

A. Segmentation Performance

If the produced images appear realistic, it is fair to assume that the expected labels are very close in compared to the traditional pictures. As a result, we use the assessment procedure used in previous work to evaluate the segmentation accuracy of the images generated Without considering the classes that can be overlooked, the mean Intersection Over-Union (mIoU) and pixel accuracy (Acc) metrics are registered. For DeepFashion, images are evaluated using well-trained segmentation models from the off-the-shelf human parser CIHP.

B. In terms of label-to-image transformation, there is a comparison.

Others, such as MulNet, GroupNet, GroupEnc, and GroupDNet, are clearly capable of producing semantically multi-modal images based on qualitative results. MulNet, GroupNet, BicycleGAN, and DSCGAN, on the other hand, have low image quality due to their implausible images. The image quality of GroupEnc is better, but it suffers in the SMIS mission. When the upper clothes are changed to a different design, the color of the short denim pants is also slightly changed by GroupEnc, as seen in the first two rows. We test their performance whether they have well-trained models available for download from their official GitHub repositories. For those experiments not included in their original reports, we follow their codes and run the experiments with the same settings as GroupDNet. On the Deep Fashion and Cityscapes datasets, our network behaves similarly to SPADE because it is based on SPADE. Although our system performs worse than SPADE on the ADE20K dataset, it still outperforms other methods.

This phenomenon shows the SPADE architecture's superiority while also exposing GroupDNet's inability to accommodate datasets with a vast number of semantic classes. In general, GroupDNet's images are more realistic and believable than those produced by others. These visual results consistently demonstrate GroupDNet's produced images' high image quality, demonstrating its effectiveness on a variety of datasets.

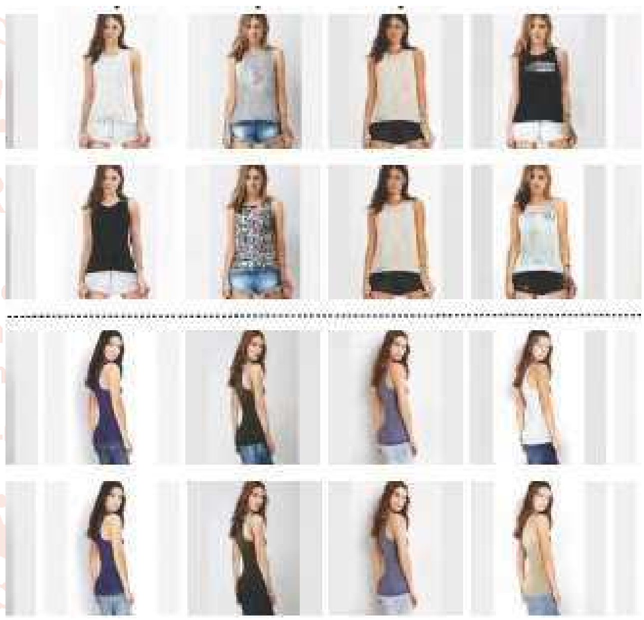

Figure 2: OutPut

V. CONCLUSION

Unlike other possible solutions such as multiple generators, our network uses all category convolution layer and adjusts the factor analysis of the convolution layers in the decoder to improve learning happens. About the fact that GroupDNet performs well on semantically multi-modal synthesis tasks and delivers relatively high-quality results, there are several problems that need to be addressed. To begin with, despite being nearly twice as fast as multiple generators networks, it takes more computing energy to practise than pix2pixHD and SPADE. Second, despite showing some minor differences in illumination, colour, and texture, for datasets with minimal variety, GroupDNet also fails to model various layouts of a single semantic type.

ACKNOWLEDGEMENTS

I should convey my real tendency and obligation to Dr MN Nachappa and Asst. Prof: Dr Kamalraj R and undertaking facilitators for their effective steerage and consistent inspirations all through my assessment work. Their ideal bearing, absolute co-action and second discernment have made my work gainful.

VI. REFERENCES

[1] "Aayush Bansal et.al, Shapes and context: In-the-wild image synthesis & manipulation. In Proc. CVPR, pages 2317–2326, 2019".

[2] "Jianmin Bao et.al, CVAE-GAN: fine-grained image generation through asymmetric training. In Proc. ICCV, pages 2764–2773, 2017".

[3] "Andrew Brock et.al, Large scale GAN training for high fidelity natural image synthesis. In Proc. ICLR, 2019'.

[4] "Qifeng Chen et.al, Photographic image synthesis with cascaded refinement networks. In Proc. ICCV, pages 1520–1529, 2017".

[5] " Marius Cordts et.al, The cityscapes dataset for semantic urban scene understanding. In Proc. CVPR, pages 3213–3223, 2016".

[6] "Vincent Dumoulin et.al, A learned representation for artistic style. In Proc. ICLR, 2017".

[7] "L. A. Gatys et.al, Image style transfer using convolutional neural networks. In Proc. CVPR, pages 2414– 2423, June 2016".

[8] "Ke Gong et.al, Instance-level human parsing via part grouping network. In Proc. ECCV, pages 805– 822, 2018".

[9] "Ian Goodfellow et.al, Generative adversarial nets. In Proc. NeurIPS, pages 2672–2680, 2014".

[10] "Shuyang Gu et.al, Mask-guided portrait editing with conditional gans. In Proc. CVPR, pages 3436–3445, 2019".